Automated Translation Between Medical Vocabularies Using a Frame-Based Interlingua

Roberto A. Rocha,M.D., Beatriz H.S.C. Rocha,M.D., Stanley M. Huff, M.D.
Department of Medical Informatics, University of Utah, Salt Lake City, UT 84112

The integration of clinical systems almost always requires a translation phase, where vocabularies are compared and the similar concepts are matched. The lack of standards in the area of medical concept representation makes this task very difficult. The authors describe the development of a frame-based application that automatically translates terms found in one vocabulary to another. The application implements an innovative scoring algorithm that ranks the best matches using an exponential scale. Preliminary results and the comparison against a manual process in the same domain are also discussed.

INTRODUCTION

The need to better utilize medical information for patient care and clinical research is constantly increasing, and the amount of information that must be analyzed and processed is also expanding. A potential solution for this problem is to develop clinical systems that can assist the analysis and use of the available information. A central problem that has retarded the development of useful clinical systems is the lack of standard methods for representing the medical terminology [1].

There are several medical vocabularies and terminologies available, but their domains and scopes are somewhat different [2]. As a result, no single medical vocabulary is accepted as a "standard" by developers or users [3]. This situation has different impact in different levels. In a hospital setting, the integration of two systems with proprietary data dictionaries is a laborious task. In a research setting, the integration of research data to literature databases is again difficult. In both cases, translation between vocabularies must be performed before the integration can be achieved.

Extensive work has been done in the area of translation between medical vocabularies [4-8]. The results are usually encouraging if you disregard the amount of human assistance needed. However, if one considers the ideal situation, where minimal human interaction is needed and extensibility to new domains is possible, further research seems to be necessary.
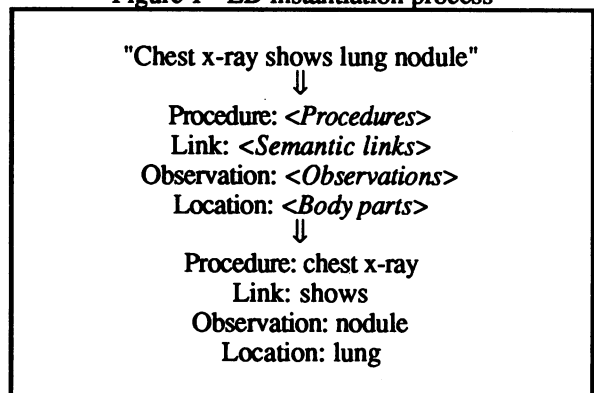
Our group has been involved in structured patient databases for over twenty-years, with most of the work concentrated on the HELP hospital information system [9]. Since the actual model being used by the HELP system does not fulfill all the needs, we have been developing alternative models for data representation that build on our previous experience. Our best current model is the Event Definition model [10,11].

The Event Definition (ED) model is based on a conceptual view that a patient's medical record is a sequence of clinical events. The ED model is a frame oriented model that resembles case frames [12]. EDs are always associated with the time that they occurred, supporting the chronological sequence of the events.

An ED captures clinical data in a semantic representation by assigning semantic meaning to the slots (attributes) in the frame (Figure 1). The semantic meaning is obtained by assigning to each slot an exclusive set of concepts that characterizes its domain. For example, a slot that represents parts of the human body has in its domain words like lung, arm, and heart. The creation and maintenance of these domains requires a lexicon, where concepts are canonically represented and identified.

Figure 1 - ED instantiation process

"Chest x-ray shows lung nodule"
⇓
Procedure: <*Procedures*>
Link: <*Semantic links*>
Observation: <*Observations*>
Location: <*Body parts*>
⇓
Procedure: chest x-ray
Link: shows
Observation: nodule
Location: lung

We will describe an application called InterMatch (Interlingua Matching), derived from our current work in the area of medical concept representation. InterMatch does not implement the full complexity of the ED model, but explores the model as an underlying structure where different controlled medical vocabularies can be represented. The convergence of different vocabularies ("languages") to an intermediate "common language" characterizes the idea of an interlingua [13]. We believe that the semantic structure of the ED model can indeed be used as an interlingua. In this sense, our approach is similar to the work developed by Masarie et al [4], although the scope and granularity of their generic frames is quite different.

METHODS

The main goal of InterMatch is to translate terms expressed in different medical vocabularies using a fully automated process. In order to demonstrate our approach, we decided to use the data dictionary of an expert system (Iliad) [14] as the "original" vocabulary, and the UMLS Metathesaurus version 1.2 (Meta) [15] as the "target" vocabulary. The domain selected for this preliminary implementation is chest radiology. Despite being a well defined domain, chest radiology clearly represents the complexity of the medical language. These choices were influenced by an ongoing parallel effort in our site that was used as comparison [16].

Building the lexicon

A necessary step before the actual use of InterMatch is to create a lexicon. The development of the lexicon started with the isolation and manual review of the terms being used by both Iliad and Meta. The objectives of the review were the assignment of a semantic type to each term and the identification (or addition) of synonymous and variant forms. In order to make the review process more accurate, for each term we also obtained a dictionary definition and isolated its context in the parent vocabulary [17].

The end product of this manual effort is a lexicon where identical concepts have the same numeric code, and all terms have a semantic type (Table 1). The terms included are as atomic as possible, i.e., representing only single concepts. For InterMatch we reused the semantic types from our previous work [11], and some adaptations of the semantic types proposed by Evans [18].

Table 1 - InterMatch lexicon (partial display)

ID	Digit	Term	Type
106	10	line shadow	651
106	11	linear shadow	651
106	20	linear opacity	651
651	10	image of pathological process	987
987	10	metalanguage term	987

The whole process described in the previous two paragraphs was in reality performed in two steps. We will give more details when we explain how we searched the target vocabulary.

Creating the ED template

In this phase, the semantic types effectively being used in the lexicon were grouped to create the slot domains. Instead of trying to characterize a single concept using a semantic type, we viewed each semantic class as a whole, combining those classes that had semantic affinities. The end result of this process is what we call an ED template. The ED template is the underlying structure that functions as an interlingua in our system.

Having created the ED template, the next step is to assign to each slot a numeric weight. The slots representing the more clinically relevant domains received the highest weights. For example, a slot that captures an observation as "opacity" has a higher precedence than a slot that captures a modifier as "right". The assignment of weights is in some sense arbitrary, but in reality this process is dictated by what is felt to be the most important concepts to match when comparing terms.

We consider the ED template as being a dynamic entity, where the slots and weights can be adjusted to prioritize different classes of concepts. This property focuses the translation according to the scope of the target vocabulary, minimizing the effect of its deficiencies.

Instantiating the ED template

The lexicon and the ED template established the content and the structure for the translation process. A heuristic parser was then used to combine both content and structure, mapping the concepts expressed in either Meta or Iliad to the actual atomic concepts stored in the lexicon. The parser acted like an automatic instantiator of the ED templates.

The output of the parser are triplets in the format of concept + semantic type + weight. Each term generated multiple triplets, one for each concept identified. Each set of triplets obtained from a single term constitutes an ED instance, and they remain linked to the original term by a unique identifier.

Searching the target vocabulary

Having described briefly the process of getting the automatic instantiations, we return to the issue of how to obtain the potential matching terms from the target vocabulary (Meta). The task of searching for potential matching terms is crucial to the success of the whole translation process. The extent and scope of Meta increased the complexity of the search process, since the domain of chest radiology was distributed among several other domains. A manual selection was clearly out of the question.

In our strategy we used the semantic structure of the ED model to guide the search. With the weights assigned to the ED template slots, we established a threshold above which the instances of the concepts

should be searched for in Meta. The weights introduced the notion of a "weighted search" where important concepts have priority over the less important ones [19,20]. In addition, the threshold helped the exclusion of those concepts that were not specific and could bring unrelated terms (functioning like a "stop list") [19].

Using these techniques, we took all the concepts that instantiated slots above a certain threshold and exploded each concept in all its variant and synonymous forms, creating a list of words. A lexicon rich in synonyms and variant forms is very important here, since we are not ultimately searching for strings but for instances of concepts. The final step was to search the occurrences of each word in the Metathesaurus word index (MRWD), obtaining a list of codes (concept unique identifiers or CUIs). The CUIs were then decoded into the actual Meta terms found in the MRCON file.

All Meta terms selected represent potential translations, and they were processed using the same steps explained before. As a result, the lexicon was expanded (new concepts) and enhanced (new synonyms and variants). New semantic types were created to accommodate some of the new concepts, and the ED template was reviewed to include the new semantic classes. All terms (Iliad and Meta) were finally parsed, generating the final instantiations.

Scoring the potential translations

With all terms from both vocabularies represented in the same underlying structure, we were ready to start the translation process. The challenge was to automatically select which term (or terms) in Meta better described the concepts being represented by a given term in Iliad. Also, in the event of a one-to-many translation, we wanted to discriminate the terms from the target vocabulary according to their clinical importance.

The challenge was solved with a scoring algorithm based on the weights assigned to the slots of the ED template. The score of an instantiated term is the summation of 2 to the power of the weight of each filled attribute. For example, if a given term has the attributes "observation" and "location" instantiated (like "heart enlargement"), and "observation" has a weight of 4 and "location" a weight of 2, the final score will be $2^4 + 2^2 = 20$.

An interesting property of this scoring algorithm assures that scores resulting from instantiations of different slots are never identical. Another interesting property is that if all the attributes below a given attribute "A" are instantiated, the resulting score is still lower than the instantiation of "A" alone (i.e.,

$2^4 > 2^0 + 2^1 + 2^2 + 2^3$). These properties ensure that the resulting exponential scale is appropriate for scoring the matches, giving to the closest match the highest score.

Another use of the scoring algorithm is to help refine the matches. Using the same example above, if the term you are trying to translate is "heart enlargement" and the potential matching term is "enlarged liver", the "observation" slot matches but the "location" does not. In this example, the location attribute counts as a negative evidence of a potential match ($2^4 - 2^2 = 12$).

Having described the scoring method, we proceed to the actual translation. We started retrieving each Iliad instantiation from the database, obtaining a group of triplets that represented their concepts. Next, taking one Iliad term (or ED instance) at a time, we isolated the concept identifiers from each triplet and retrieved all Meta instantiations that contained those same concept identifiers. Each Meta instantiation was scored. If the resulting score was above zero, the term was retained. After processing all Meta instantiations for a particular Iliad term, we sorted all retained terms in descending order. Those terms with the highest scores in each range (range defined as the interval between two slots) represented the closest matching terms (Table 2). Calculating the same score for the Iliad term being translated, we were able to compare how close the Meta terms were from an exact match.

Table 2 - Example of the results obtained with the scoring method

Original Term	Score	Target Term
Airway Obstruction	320	Obstructions, Airway
Airway Obstruction	256	Obstruction
Airway Obstruction	192	Nasal Obstruction
Airway Obstruction	64	Airways
Airway Obstruction	-192	Airway Resistance

Validating the method

As mentioned before, there is an ongoing effort in our site to translate the Iliad data dictionary into Meta [16]. This research project provided the opportunity to compare the performance of InterMatch to a manual method. The main goal was to verify if InterMatch was at least comparable to the manual review performed by a physician. In addition, we analyzed how complete the output from InterMatch was in terms of finding the potential terms, as well as how accurate the scoring algorithm was in ranking the best potential match.

RESULTS

The lexicon used by InterMatch has 4351 entries, representing 3070 concepts (average of 1.42 terms per concept). The time spent reviewing each term of the lexicon ranged from 5 to 60 seconds, with an average of 9 seconds per term. The concepts represented in the lexicon were subdivided in 65 semantic classes that were linked to an ED template with 38 slots.

From Iliad we isolated 150 terms describing the chest x-ray domain. Parsing these terms we obtained 1041 filled slots, representing an average of 6.94 slots per ED instance. The concepts represented in these 150 Iliad terms retrieved 2727 Meta terms. The parser generated 8174 filled slots from these Meta terms, with and average of 3.0 slots per ED instance.

The automated process and the manual process were compared using McNemar's test for correlated proportions (normal theory method). Table 3 shows the counts of concordant and discordant pairs for the 150 cases. The Chi-Square obtained was 0.3556 with p equal to 0.5510.

Table 3 - Counts of concordant and discordant pairs.

		Manual match	no match
InterMatch	match	47	20
	no match	25	58

We also analyzed how InterMatch performed in identifying the concepts necessary for the translation, (Table 4). In addition, we analyzed the scoring algorithm performance in assigning the highest score to the best match (Table 5).

Table 4 - Concepts identified

	Count	%
Both	69	49.60
InterMatch	39	28.10
Manual	31	22.30

Table 5 - Scoring algorithm performance.

Best match position	Count	Cumulative %
1st	57	70.00
2nd	20	94.00
3rd	5	100.00

DISCUSSION

The manual effort to create the lexicon for InterMatch is certainly the most laborious phase of our approach. The lexicon is the core of the whole system and the success of the several routines involved depend on it. Nevertheless, the effort to create a lexicon is not in vain, since one can reuse and expand it as needed. In fact, the expansions tend to become smaller as the coverage of the domain improves.

Several authors advocate the usefulness of a semantically typed lexicon [2,3,5,6,17], and some recent developments in our area indicate that sharing some of this knowledge is indeed a necessity. These previous comments are somewhat true for the frame-based model we have been using. Other research groups use similar data models, showing that the theory involved is not complex [3-5,21].

The comparison between InterMatch and the manual translation fulfilled our expectations. The difference in the performance of both systems is not statistically significant (no significant disagreement). In other words, in this limited domain InterMatch can produce translations that are comparable to a manual review performed by a physician. The methodology used by InterMatch was always the same, improving dramatically the consistency of the final product. However, the effort to create the lexicon certainly surpassed the effort spent to do the manual translation. The development of a lexicon is likely to justify itself only if one needs to translate a much larger set of terms. This conclusion does not invalidate our approach, since we needed a pilot study for testing our application.

InterMatch was unable to identify 22.30% of the total number of concepts matched (Table 4). The analysis of these particular terms reveals basically deficiencies in the lexicon that can be easily fixed. These deficiencies include lack of synonyms, incompatible term granularity, and semantic misclassification. We expect to learn how to avoid this problems as we progress. A good sign was the 39 (28.05%) additional concepts that only InterMatch was able to identify (Table 4).

The scoring method also produced satisfactory results, being able to bring to the top the best match in 70% of the cases (Table 5). In fact, all the best matches were found in the first three highest scores (Table 5). The performance was not ideal, probably because of the deficiencies in the lexicon. In addition, this scoring method becomes less precise when a term ends up having the same slot instantiated twice by different concepts, i.e., "acute or chronic gastritis". These "collisions" did not seem to affect very much our results, but we are working on a routine that will handle this problem.

The overall implementation of InterMatch and its comparison to the manual method identified the potential for several enhancements. We believe we

can improve the performance if we make use of hierarchical inference, meaning that broader or narrower terms will also become possible matches. We also plan to explore a link to the UMLS semantic network to help handle ambiguous terms. Finally, we may enhance the parser algorithm with a formal grammar, especially to be able to handle negation and uncertainty.

One of the greatest challenges for the next years will be the functional integration of the available information system [21]. In this scenario, any application that can speed up the laborious task of matching terminologies used by different systems is very useful. Our results demonstrate that InterMatch has a potential to become one of these applications.

Reference

[1] Blois MS. Information and Medicine - The Nature of Medical Descriptions. University of California Press, 1984.

[2] Rossi Mori A, Bernauer J, Pakarinen V, Rector AI, Robbè P, Ceusters W, Hurlen P, Ogonowski A, Olesen H. Model for Representation of Terminologies and Coding Systems in Medicine. *Proc. the Seminar "Opportunities for European and U.S. Cooperation in Standardization in Health Care Informatics"*, Geneva, September 1992.

[3] Cimino JJ, Hripcsak G, Johnson SB, Friedman C, Fink DJ, Clayton PD. UMLS as a Knowledge Base - A Rule-Based Expert System Approach to Controlled Medical Vocabulary Management. *Proc. of the 14th Symposium on Comp. Applic. in Medical Care*, 175-179, 1990.

[4] Masarie FE, Miller RA, Bouhaddou O, Giuse NB, Warner, HR. An Interlingua for Electronic Interchange of Medical Information: Using Frames to Map between Clinical Vocabularies. *Computers and Biomedical Research* 24: 379-400, 1991.

[5] Evans DA, Rothwell DJ, Monarch IA, Lefferts RG, Côté RA. Toward Representations for Medical Concepts. *Medical Decision Making* 11(4): S102-S108 (supplement), 1991.

[6] Cimino JJ, Barnett GO. Automated Translation Between Medical Terminologies Using Semantic Definitions. *MD Computing* 7(2): 104-109, 1990.

[7] Wingert F. Medical Linguistics: Automated Indexing into SNOMED. *CRC Critical Reviews in Medical Informatics* 1(4): 335-403, 1987.

[8] Sherertz DD, Tuttle MS, Blois MS, Erlbaum MS. Intervocabulary Mapping within the UMLS: The Role of Lexical Matching. *Proc. of the 12th Symposium on Comp. Applic. in Medical Care*, 201-206, 1988.

[9] Kuperman GJ, Gardner RM, Pryor TA. HELP: A Dynamic Hospital Information System. Springer-Verlag, 1991.

[10] Huff SM, Craig RB, Gould BL, Castagno DL, Smilan RE. Medical Data Dictionary for Decision Support Applications. *Proc. of the 11th Symposium on Comp. Applic. in Medical Care*, 310-317, 1987.

[11] Rocha RA, Huff SM, Haug PJ, Koehler S, Wang P. Towards a Representation Scheme for Clinical Data. *Abstract accepted for the 1993 AMIA Spring Congress*, St. Louis, May 9-12.

[12] Schank RC, Colby KM, eds. Comp. Models of Thought and Language. WH Freeman, 1973.

[13] Weaver, W. Translation. In Locke and Booth, 15-23, 1955.

[14] Warner HR, Haug P, Lincoln M, Warner H Jr, Sorenson D, Fan C. Iliad as an Expert Consultant to Teach Differential Diagnosis. *Proc. of the 12th Symposium on Comp. Applic. in Medical Care*, 371-376, 1988.

[15] Tuttle MS, Blois MS, Erlbaum MS, Stuart JN, Sherertz DD. Toward a Bio-medical Thesaurus: Building the Foundation of the UMLS. *Proc. of the 12th Symposium on Comp. Applic. in Medical Care*, 191-195, 1988.

[16] Bouhaddou O, Warner H, Huff S, Bray B. Sorenson D, Dougherty N. Evaluating How the UML Meta1.1 Covers Disease Information Contained in a Diagnostic Expert System (Iliad). *Abstract accepted for the 1993 AMIA Spring Congress*, St. Louis, May 9-12.

[17] Evens MW ed. Relational models of the lexicon - Representing knowledge in semantic networks. Cambridge University Press, 1988.

[18] Evans DA. The MedSort II Project: Developing and Managing Medical Thesauri. Final Report Contract No. No1-LM-6-3512, National Library of Medicine, 1987.

[19] Frakes WB, Baeza-Yates R, eds. Information Retrieval - Data Structures & Algorithms. Prentice Hall, 1992.

[20] Hersh WR, Greenes RA. Information Retrieval in Medicine: State of the Art. *MD Computing*, 7(5): 302-311, 1990.

[21] Linnarsson R, Wigertz O.. The Data Dictionary - A Controlled Vocabulary for Integrating Clinical Databases and Medical Knowledge Bases. *Methods of Information in Medicine*, 28(2): 78-85, 1989.